

# Documenting data reporting formats on GitHub



**Rob Crystal-Ornelas**

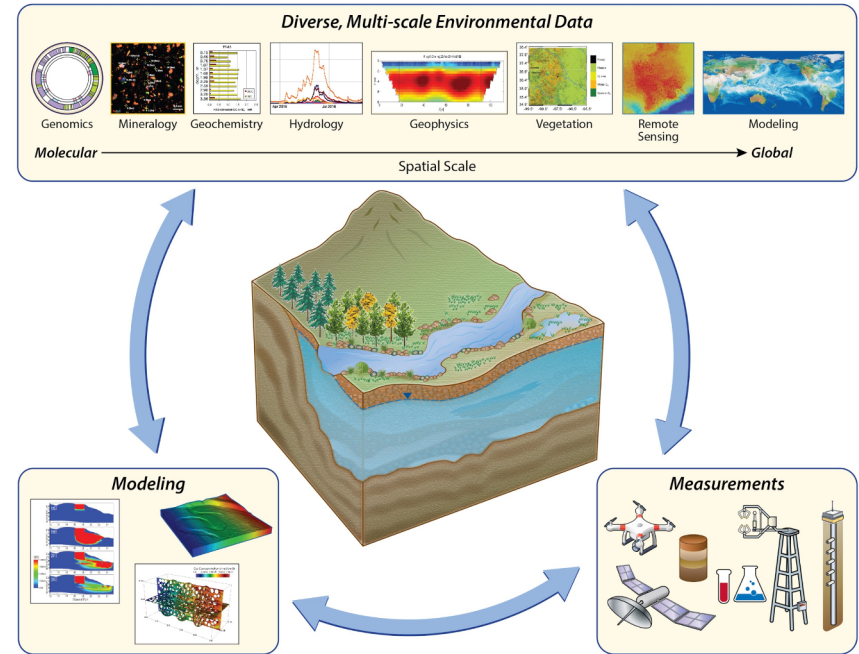
Deb Agarwal, Charuleka Varadharajan, Shreyas Cholia,  
Valerie Hendrix, Joan Damerow, Fianna O'Brien, Hesham Elbashandy,  
Mario Melara, Makayla Shepherd, Bill Riley, Madison Burrus,  
Zarine Kakalia, Emily Robles, Maegen Simmonds, Cory Snively, and  
Karen Whitenack



ESS-DIVE Sept 2020 webinar

# ESS-DIVE: A DOE cross-disciplinary data repo

- **Preserve**, expand **access to**, and improve **usability** of **data** generated through DOE-sponsored research.
- **Terrestrial and subsurface ecosystems data**
- Support DOE's efforts to address pressing energy and environmental challenges.



# Presentation overview

- Data reporting formats and GitHub
- **Systematic review** of how organizations manage data reporting formats
- **ESS-DIVE's Community Space on GitHub**

# Data reporting formats and GitHub



# Valuable data are being lost with time

> [Nature](#). 2020 Aug;584(7822):656-658. doi: 10.1038/d41586-020-02462-7.

## **Challenge to scientists: does your ten-year-old code still run?**

[Jeffrey M Perkel](#)

PMID: 32839567 DOI: [10.1038/d41586-020-02462-7](#)

# Valuable data are being lost with time

> [Nature](#). 2020 Aug;584(7822):656-658. doi: 10.1038/d41586-020-02462-7.

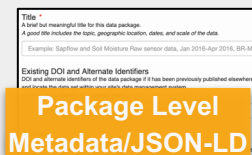
**Challenge to scientists: does your ten-year-old code still run?**

Jeffrey M Perkel

PMID: 32839567 DOI: [10.1038/d41586-020-02462-7](#)

## Can you find & use your ten-year-old data?

# Reporting format efforts to make diverse ESS data FAIR



Agarwal, Hendrix (LBNL)

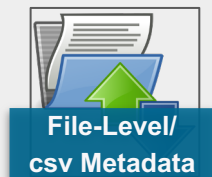


Damerow (LBNL)



Damerow (LBNL)

ESS-DIVE team developed **high priority, base data formats**



Velliquette, Heinz,  
Devarakonda (ORNL)



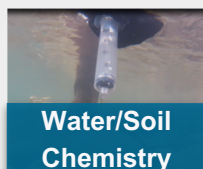
Bond-Lamberty,  
Pennington (PNNL)



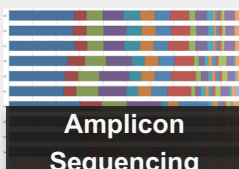
Rogers, Ely (BNL)



Goldman (PNNL)



Boye (SLAC)

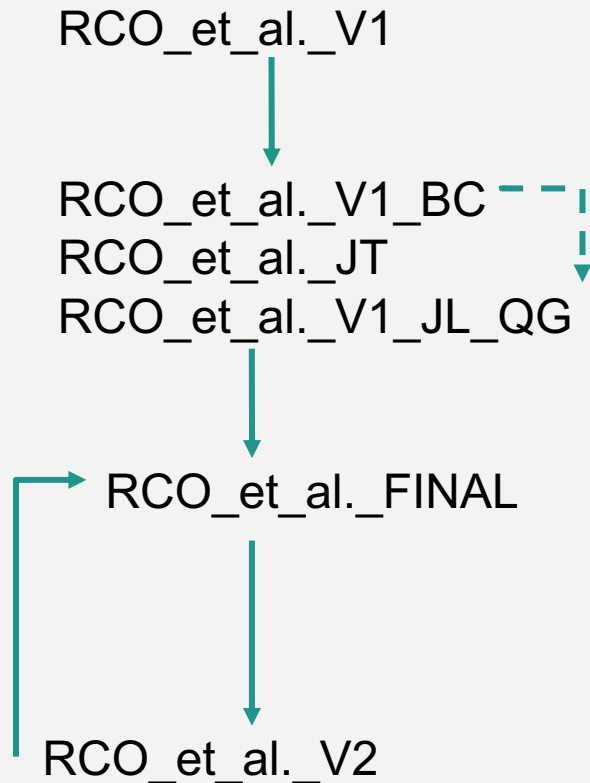


Weisenhorn (ANL)

Developing **reporting formats for select ESS data types** using \$1M community funds

# GitHub as a place to develop reporting formats

Not very efficient



“Like track changes [...] but more rigorous, powerful, and **scaled up to multiple files**.”  
(Bryan, 2018)

## Version control

Motivation for our work:

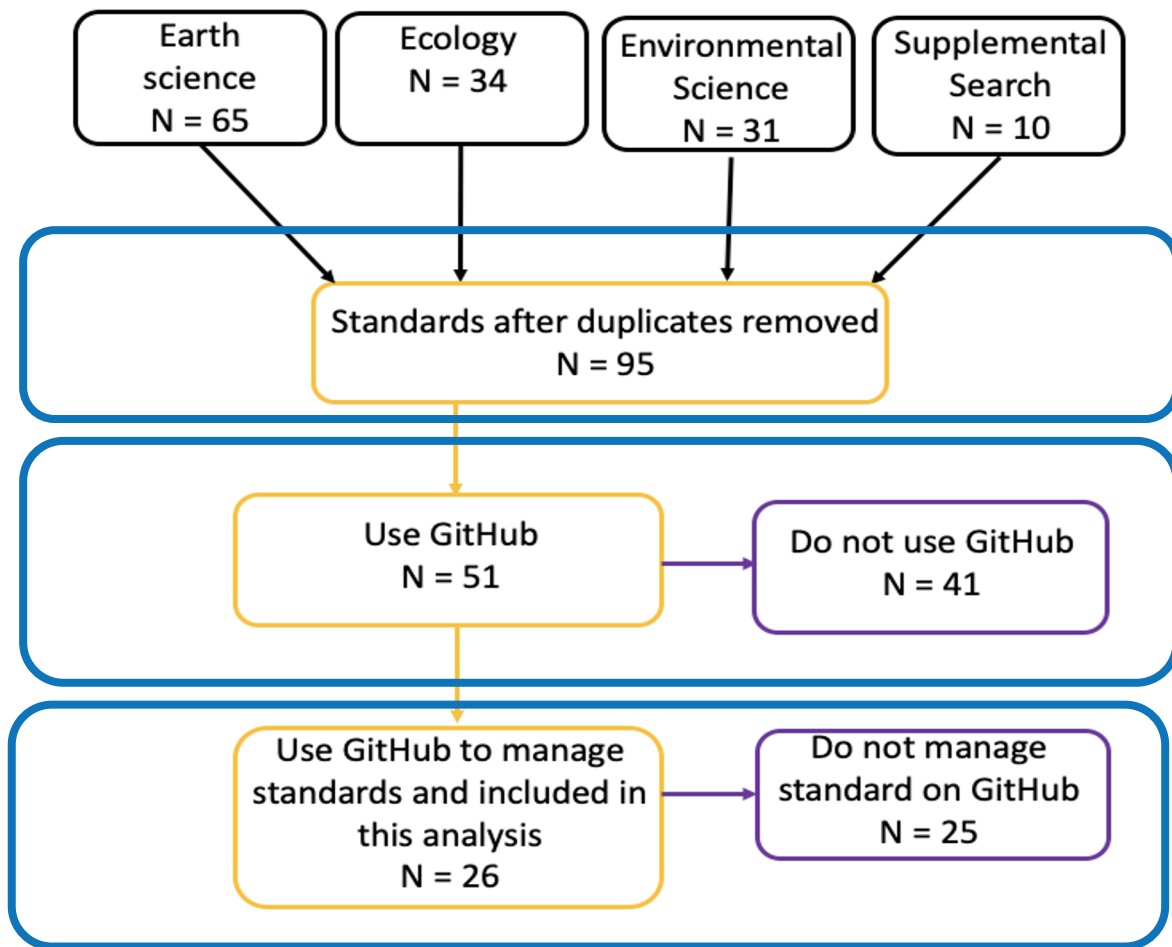
1. Community oriented--feedback from users
2. Open versioning
3. User-friendly documentation

# **Systematic Review:**

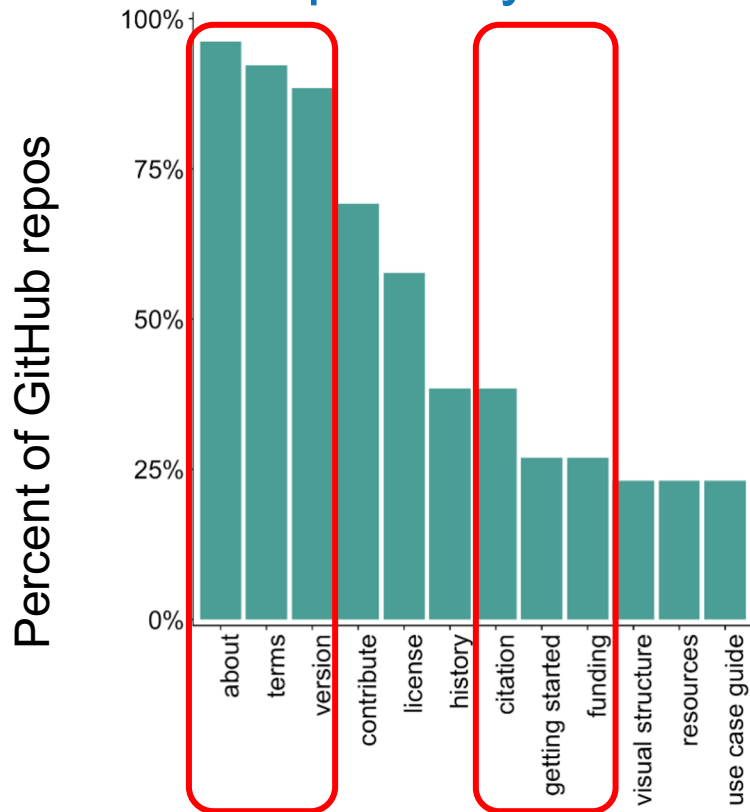
How organizations manage data reporting formats

Many groups use  
GitHub for version  
control of code **and**  
**text-based**  
**documents**

Darwin Core

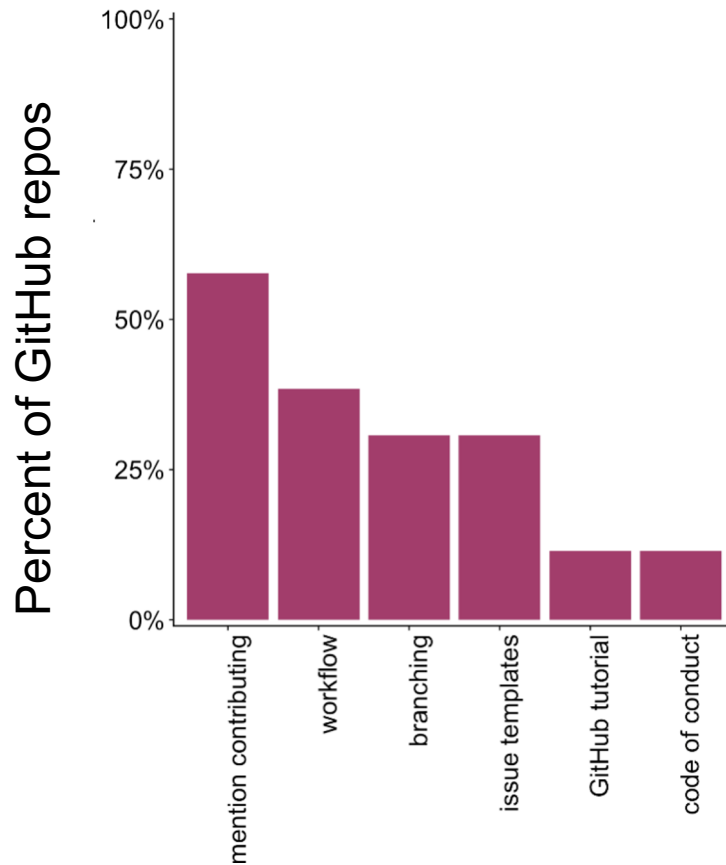


## GitHub repository content



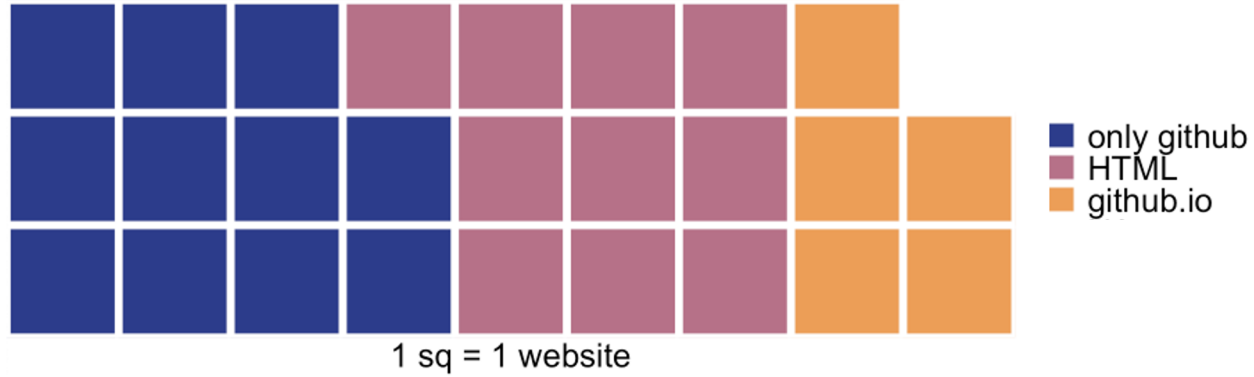
GitHub repo elements

## How data users contribute



Contributing elements

# User facing web presence





## What we found in the systematic review:

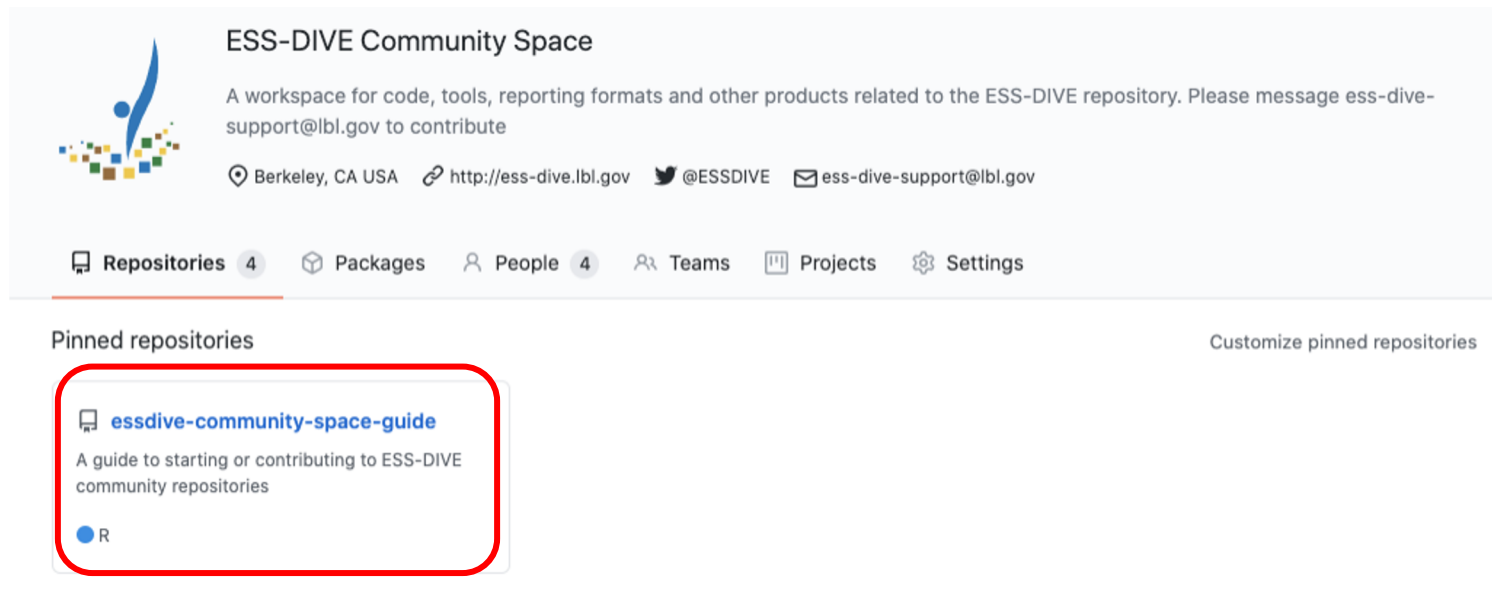
- Wide variety in what groups decide to display on GitHub
- ~50% mention contributing, but only 33% provide guidance
- User facing site is often GitHub repository

# ESS-DIVE & GitHub:

Using GitHub to manage ESS-DIVE's suite of data reporting formats

# ESS-DIVE's GitHub community space

<https://github.com/ess-dive-community>



The screenshot shows the GitHub profile page for the ESS-DIVE Community Space. At the top, there is a profile picture of a stylized blue figure with colorful dots, the name "ESS-DIVE Community Space", and a bio: "A workspace for code, tools, reporting formats and other products related to the ESS-DIVE repository. Please message ess-dive-support@lbl.gov to contribute". Below the bio are links for location (Berkeley, CA USA), website (http://ess-dive.lbl.gov), Twitter (@ESSDIVE), and email (ess-dive-support@lbl.gov). A navigation bar includes "Repositories 4", "Packages", "People 4", "Teams", "Projects", and "Settings". The "Pinned repositories" section is visible, with the first repository, "essdive-community-space-guide", highlighted by a red rounded rectangle. This repository is described as "A guide to starting or contributing to ESS-DIVE community repositories" and is associated with the R programming language.


**ESS-DIVE Community Space**


A workspace for code, tools, reporting formats and other products related to the ESS-DIVE repository. Please message [ess-dive-support@lbl.gov](mailto:ess-dive-support@lbl.gov) to contribute

📍 Berkeley, CA USA 🔗 <http://ess-dive.lbl.gov> 🐦 @ESSDIVE ✉ [ess-dive-support@lbl.gov](mailto:ess-dive-support@lbl.gov)

📁 Repositories 4 📦 Packages 👤 People 4 👥 Teams 📁 Projects ⚙ Settings

**Pinned repositories** [Customize pinned repositories](#)

 **[essdive-community-space-guide](#)**  
A guide to starting or contributing to ESS-DIVE community repositories  
● R

 robcrystalornelas	Update README.md	4c98d04 1 minute ago	🕒 40 commits
📁 images	Added new images and new text to starting repo readme		last month
📄 LICENSE.md	Create LICENSE.md		18 days ago
📄 README.md	Update README.md		1 minute ago
📄 csv_to_markdown.R	modified this repo so that it is now one single readme file with subs...		last month
📄 template_for_README.md	updating readme text		5 minutes ago

README.md

## Guide to the ESS-DIVE Community Space



“If you can write an email, you can write in markdown” (Bryan, 2018)

## The purpose of the Community Space

---

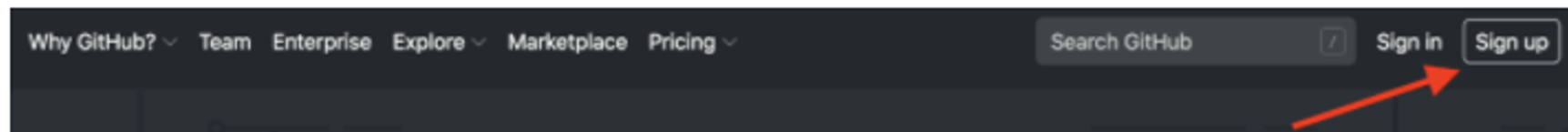
The ESS-DIVE GitHub Community Space was started to provide a location where the ESS-DIVE user community can upload documents, codes, and tools related to the [ESS-DIVE repository](#). We designed this meta-repository so that ESS-DIVE contributors can familiarize themselves with the process of contributing to an existing GitHub repository or starting their own.

## How to contribute

---

### Setting up a GitHub account

The first step for anyone interested in contributing to the ESS-DIVE GitHub community space is to go to the [GitHub website](#) and create a GitHub account or sign in if you already have an account.





robcrystalornelas Update README.md

4c98d04 1 minute ago 🕒 40 commits

📁 images	Added new images and new text to starting repo readme	last month
📄 LICENSE.md	Create LICENSE.md	18 days ago
📄 README.md	Update README.md	1 minute ago
📄 csv_to_markdown.R	modified this repo so that it is now one single readme file with subs...	last month
📄 template_for_README.md	updating readme text	5 minutes ago

README.md



## Guide to the ESS-DIVE Community Space






## Template for Community Space README files

---

In this document, you will find our suggested template for creating a GitHub repository README.md file. Readme files are written in the markdown language and when they are rendered on GitHub or in a GitBook, they are easily readable as a webpage. This site provide a helpful guide for the type of coding that can be done in markdown: <https://commonmark.org/help/>

We provide the following suggestions based on a review that we conducted of 26 GitHub repositories that use this platform for managing both text-based data reporting format documentation.

---



## Getting started

---

The getting started section of your repository is like a table of contents that lists out all of the content a reader can expect to find in your repository. You do not need to link to every single document in your GitHub repo, but you should link to the most essential documents, so that visitors can access them quickly.

---



## How to contribute

---

Provide details here to let readers know how to suggest modifications to your repository. We recommend that you give a direct link to your [GitHub issues page](#). You may also choose to provide contributing templates, so that when users submit a GitHub issue, they do so in a format you specify. Examples of issue templates can be found [here](#)

---



## About the [reporting format, computer code, tool]

---

# The benefits of using GitHub “issues”

 **ess-dive-community / essdive-sample-id-metadata**

 Code

 **Issues** 2

 Pull requests

 Actions

 Projects



robcrystalornelas commented now

Member

Author



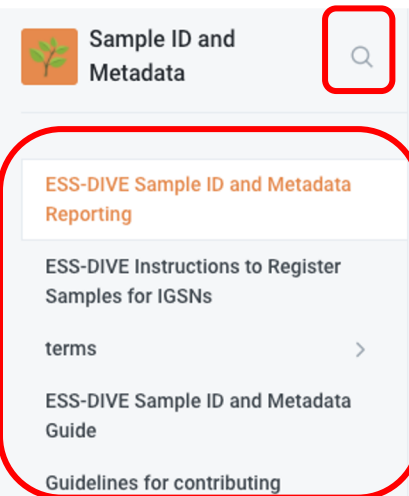
## Change documentation

- Documentation reference (URL): <https://github.com/ess-dive-community/essdive-sample-id-metadata/blob/master/contribute.md>
- Submitter: Rob Crystal-Ornelas
- Justification (why is this change necessary?): renaming one of our files so that it adheres to commonly used naming conventions.

I suggest the following changes: Renaming contribute.md to CONTRIBUTING.md The new file name will be consistent with how other groups that use github label their file about contributing to repos.



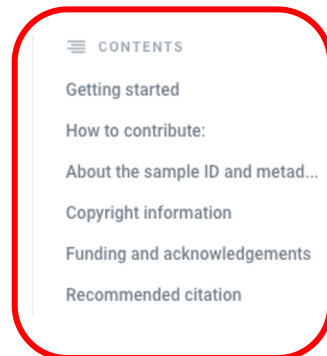
# Easy translation of GitHub content into user-friendly documentation



## ESS-DIVE Sample ID and Metadata Reporting

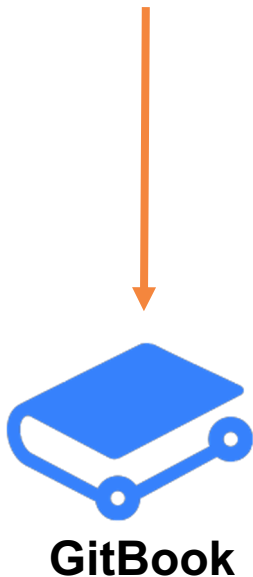
ESS-DIVE recommends registering samples for **Global Sample Numbers (IGSNs)** through the **System for Earth Sample Registration (SESAR)**. IGSNs are associated with standardized metadata to characterize a variety of different samples and their collection details. These sample identifiers facilitate sample discovery, tracking, and reuse; they are especially useful when sample data is shared with collaborators, sent to different labs or user facilities for analyses, or distributed in different data files, datasets, and/or publications.

ESS-DIVE has worked with our community scientists to test use of IGSNs and associated



<https://ess-dive.gitbook.io/sample-id-and-metadata/>

# Updates to GitHub instantly pushed to the GitBook



<> Edit file

👁 Preview changes

```
1  # ESS-DIVE Sample ID and Metadata Reporting
2
3  ESS-DIVE recommends registering samples for [Global Sample Numbers
```



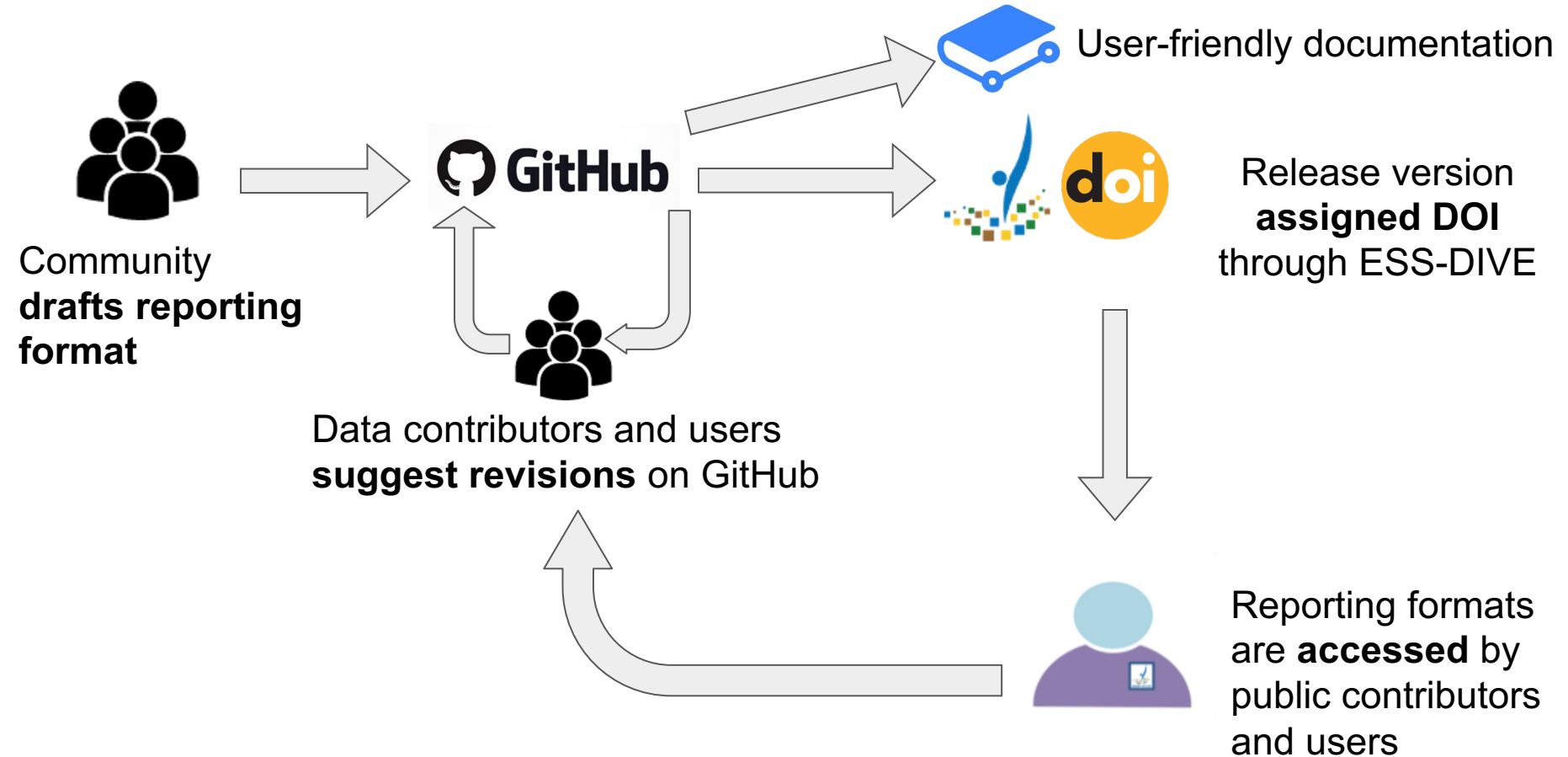
Sample ID and  
Metadata



ESS-DIVE Sample ID and Metadata  
Reporting

## ESS-DIVE Sample ID and Metadata Reporting

# Workflow for publishing the data format



# Summary

- GitHub is an agile platform for managing reporting formats:  
issues, changes all transparent
- GitBooks are a user-friendly way to view GitHub repo

# Thank you!



- ESS-DIVE teams
- Community partners



rcrystalornelas@lbl.gov



ESS-DIVE Sept 2020 webinar